Amendments to the Claims:

This listing of claims will replace all prior versions of the claims.

1. (Currently Amended) A process to evaluate an input string to segment said string into component parts comprising:

providing a state transition model derived from training data from an existing collection of data records that includes probabilities to segment input strings into component parts, ~~which~~ wherein the training data corresponding to database attributes in the existing collection of data records does not comprise manually segmented training data, and the state transition model categorizes tokens in database attribute values of the data records into positions based on a fixed beginning, middle, and trailing token topology that:

categorizes each boundary token of a database attribute value that includes multiple tokens into corresponding beginning and trailing positions,

categorizes each token that does not comprise a boundary token of a database attribute value into a middle position,

defines beginning, middle, and trailing state categories, wherein each state category includes ~~categorizes~~ states ~~for accepting classes of~~ that accept tokens ~~into~~ only if appearing in a corresponding one of said beginning, middle, and trailing positions, and

adjusts said states and probabilities associated with said states within said ~~positions~~ state categories in order to relax sequential specificity and account for erroneous token placement when evaluating tokens in the input string appearing in particular positions, wherein the state category corresponding to a particular position in which the token appears is adjusted to include states from another state category that accept tokens appearing in a different position~~, wherein training data corresponding to database attributes in the existing collection of data records does not comprise manually segmented training data~~;

determining a most probable segmentation of the input string by comparing tokens that make up the input string with the state transition model derived from the existing collection of data records;

segmenting the input string into one or more component parts according to the most probable segmentation; and

storing the one or more component parts in a database on a computer system.

2. (Original) The process of claim 1 wherein the state transition model has probabilities for multiple states of said model and a most probable segmentation is determined based on a most probable token emission path through different states of the state transition model from a beginning state to an end state.

3. (Previously Presented) The process of claim 1 wherein the collection of data records is stored in a database relation and an order of attributes for the database relation as the most probable segmentation is determined.

4. (Original) The process of claim 3 wherein the input string is segmented into sub-components which correspond to attributes of the database relation.

5. (Previously Presented) The process of claim 4 wherein the tokens that make up the input string are substrings of said input string.

6. (Original) The process of claim 5 wherein the input string is to be segmented into database attributes and wherein each attribute has a state transition model based on the contents of the database relation.

7. (Original) The process of claim 6 wherein the state transition model has multiple states for a beginning, middle, and trailing position within an input string.

8. (Original) The process of claim 6 wherein the state transition model has probabilities for the states and a most probable segmentation is determined based on a most probable token emission path through different states of the state transition model from a beginning state to an end state.

9. (Original) The process of claim 5 wherein input attribute order for records to be segmented is known in advance of segmentation of an input string.

10. (Previously Presented) The process of claim 5 wherein an attribute order is learned from a batch of records that are inserted into the state transition model.

11. (Original) The process of claim 6 wherein the state transition model has at least some states corresponding to base tokens occurring in the reference relation.

12. (Original) The process of claim 6 wherein the state transition model has class states corresponding to token patterns within said reference relation.

13. (Previously Presented) The process of claim 8 wherein the state transition model includes states that account for missing, misordered and inserted tokens within an attribute.

14. (Currently Amended) The process of claim 13 wherein:

the state transition model ~~has a beginning, a middle and a trailing state topology and the process of accounting~~ accounts for misordered and inserted tokens ~~is performed~~ when evaluating a boundary token in the input string by copying states from ~~one of said beginning,~~ a middle state category ~~or trailing states~~ into ~~another of said~~ one of a beginning state category~~, middle~~ or a trailing ~~states~~ state category, and

the state transition model accounts for misordered and inserted tokens when evaluating a middle token in the input string by copying states from the beginning state category or the trailing state category into the middle state category.

15. (Currently Amended) A computer readable storage medium containing instructions ~~that when executed cause~~ causing a computer to perform the process of claim 1.

16. (Currently Amended) A process for segmenting strings into component parts comprising:

providing a reference table of string records that are segmented into multiple substrings corresponding to database attributes, wherein the reference table of string records does not comprise manually segmented training data;

analyzing the substrings within database attribute values of string records for an attribute during a training phase to provide a state model that categorizes the substrings within database attribute values into positions based on a fixed beginning, ~~a~~ middle, and ~~a~~ trailing token topology for said attribute~~, said topology including~~ that:

categorizes beginning and trailing substrings of a database attribute value that includes multiple substrings into corresponding beginning and trailing positions,

categorizes each substring of the database attribute value that does not comprise a beginning or trailing substring into a middle position,

accepts a null token for an empty attribute component[[;]],

defines beginning, middle, and trailing state categories, wherein each state category includes ~~categorizing~~ states ~~for accepting classes of~~ that accept tokens ~~into~~ only if appearing in a corresponding one of said beginning, middle, and trailing positions, and

adjusts said states and probabilities associated with said states within said state categories in order to relax sequential specificity and account for erroneous token placement when evaluating tokens in the input string appearing in particular positions, wherein the state category corresponding to a particular position in which the token appears is adjusted to include states from another state category that accept tokens appearing in a different position;

breaking an input record into a sequence of tokens;

determining a most probable segmentation of the input record by comparing the tokens of the input record with state models derived for attributes from the reference table;

segmenting the input record into one or more component parts according to the most probable segmentation; and

storing the one or more component parts in a database on a computer system.

17. (Currently Amended) A computer system for processing input strings to segment those records for inclusion into a database comprising:

a) a database management system to store records organized into relations wherein data records within a relation are organized into a number of attributes;

b) a model building component on the computer system that builds a number of attribute recognition models derived from training data from an existing relation of data records, wherein training data corresponding to database attributes in the existing relation of data records does not comprise manually segmented training data, wherein one or more of said attribute recognition models includes probabilities for segmenting input strings into component parts which categorizes tokens in database attribute values of the data records into positions based on a fixed beginning, middle, and trailing token topology that:

categorizes each boundary token of a database attribute value that includes multiple tokens into corresponding beginning and trailing positions,

categorizes each token that does not comprise a boundary token of a database attribute value into a middle position,

defines beginning, middle, and trailing state categories, wherein each state category includes ~~categorizes~~ states ~~for accepting classes of~~ that accept tokens ~~into~~ only if appearing in a corresponding one of said beginning, middle, and trailing positions, and

adjusts said states and probabilities associated with said states within said ~~positions~~ state categories in order to relax sequential specificity and account for erroneous entries when evaluating tokens within an input string appearing in particular positions, wherein the state category corresponding to a particular position in which the token appears is adjusted to include states from another state category that accept tokens appearing in a different position; and

c) a segmenting component on the computer system that receives an input string and determines a most probable record segmentation by evaluating transition probabilities of states within the attribute recognition models built by the model building component.

18. (Original) The system of claim 17 wherein the segmenting component receives a batch of evaluation strings and determines an attribute order of strings in said batch and thereafter assumes the input string has tokens in the same attribute order as the evaluation strings.

19. (Original) The system of claim 18 wherein the segmenting component evaluates the tokens in an order in which they are contained in the input string and considers state transitions from multiple attribute recognition models to find a maximum probability for the state of a token to provide a maximum probability for each token in said input string.

20. (Currently Amended) The system of claim 17 wherein:

the model building component ~~assigns states for each attribute for a beginning, middle and trailing token position and wherein the model building component relaxes token acceptance by the model~~ accounts for erroneous entries when evaluating a boundary token in the input string by copying states ~~among said beginning,~~ from a middle state category into one of a beginning state category and a trailing ~~token positions~~ state category, and

the model building component accounts for erroneous entries when evaluating a middle token in the input string by copying states from the beginning state category or the trailing state category into the middle state category.

21. (Original) The system of claim 20 wherein the model building component defines a start and end state for each model and accommodates missing attributes by assigning a probability for a transition from the start to the end state.

22-24. (Cancelled)

25. (Previously Presented) The process of claim 1 wherein determining a most probable segmentation of the input string comprises:

considering a first token in the input string and determining a maximum state probability for said first token based on state transition models for multiple data table attributes, and

considering in turn subsequent tokens in the input string and determining maximum state probabilities for said subsequent tokens from a previous token state until all tokens are considered, and

wherein segmenting the input string comprises segmenting the input string by assigning the tokens of the input string to attribute states of the state transition models corresponding to said maximum state probabilities, wherein the state transition models are derived from training data from the existing collection of data records that does not comprise manually segmented training data.

26. (Previously Presented) The process of claim 25 further comprising determining an attribute order for a batch of input string records and using the order to limit the possible state probabilities when evaluating tokens in the input string.

27. (Currently Amended) A system for evaluating an input string to segment said input string into component parts comprising:

means for providing a state transition model derived from training data from an existing collection of data records that includes probabilities to segment input strings into component parts, ~~which~~ wherein the training data corresponding to database attributes in the existing collection of data records does not comprise manually segmented training data, and the state transition model categorizes tokens in database attribute values of the data records into positions based on a fixed beginning, middle, and trailing token topology that:

categorizes each boundary token of a database attribute value that includes multiple tokens into corresponding beginning and trailing positions,

categorizes each token that does not comprise a boundary token of a database attribute value into a middle position,

defines beginning, middle, and trailing state categories, wherein each state category includes ~~categorizes~~ states ~~for accepting classes of~~ that accept tokens ~~into~~ only if appearing in a corresponding one of said beginning, middle, and trailing positions, and

adjusts said states and probabilities associated with said states within said ~~positions~~ state categories in order to relax sequential specificity and account for erroneous token placement when evaluating tokens in the input string appearing in particular positions, wherein the state category corresponding to a particular position in which the token appears is adjusted to include states from another state category that accept tokens appearing in a different position~~, wherein training data corresponding to database attributes in the existing collection of data records does not comprise manually segmented training data~~;

means for determining a most probable segmentation of the input string by comparing an order of tokens that make up the input string with the state transition model derived from the existing collection of data records;

means for segmenting the input string into one or more component parts according to the most probable segmentation; and

means for storing the one or more component parts in a database on a computer system.

28. (Original) The system of claim 27 wherein the state transition model has probabilities for multiple states of said model and a most probable segmentation is determined based on a most probable token emission path through different states of the state transition model from a beginning state to an end state.

29. (Previously Presented) The system of claim 27 additionally including means for maintaining a collection of records, wherein the collection of records is stored in a database relation.

30. (Previously Presented) The system of claim 29 wherein the input string is segmented into sub-components which correspond to attributes of the database relation.

31. (Previously Presented) The system of claim 30 wherein the tokens that make up the input string are substrings of said input string.

32. (Original) The system of claim 30 wherein the input string is to be segmented into database attributes and wherein each attribute has a state transition model based on the contents of the database relation.

33. (Currently Amended) The system of claim 32 wherein each state category of the state transition model has multiple states for accepting a ~~beginning,~~ boundary token or a middle ~~and trailing position~~ token within an input string.

34. (Original) The system of claim 32 wherein the state transition model has probabilities for the states and a most probable segmentation is determined based on a most probable state path through different states of the state transition model from a beginning state to an end state.